DOI: 10.3745/KIPSTB.2011.18B.1.021

얼굴 깊이 추정을 이용한 3차원 얼굴 생성 및 추적 방법

주 명 호[†] · 강 행 봉[††]

요 약

얼굴의 3차원 정보는 얼굴 인식이나 얼굴 합성, Human Computer Interaction (HCI) 등 다양한 분야에서 유용하게 이용될 수 있다. 그러나 일반적으로 3차원 정보는 3D 스캐너와 같은 고가의 장비를 이용하여 획득되기 때문에 얼굴의 3차원 정보를 얻기 위해서는 많은 비용이 요구된다. 본 논문에서는 일반적으로 손쉽게 얻을 수 있는 2차원의 얼굴 영상 시퀀스로부터 효과적으로 3차월 얼굴 형태를 추적하고 재구성하기 위한 3차원 Active Appearance Model (3D-AAM) 방법을 제안한다. 얼굴의 3차원 변화 정보를 추정하기 위해 학습 영상은 정면 얼굴 포즈로 다양한 얼굴 표정 변화를 포함한 영상과 표정 변화를 갖지 않으면서 서로 크게 다른 얼굴 포즈를 갖는 영상으로 구성한다. 입력 영상의 3차원 얼굴 변화를 추정하기 위해 먼저 서로 다른 포즈를 갖는 학습 영상으로부터 얼굴의 각 특징점(Land-mark)의 기하학적 변화를 이용하여 깊이 정보를 추정하고 추정된 특징점의 깊이 정보를 입력 영상의 2차원 얼굴 변화에 추가하여 최종적으로 입력 얼굴의 3차원 변화를 추정한다. 본 논문에서 제안된 방법은 얼굴의 다양한 표정 변화와 함께 3차원의 얼굴 포즈 변화를 포함한 실험 영상을 이용하여 기존의 AAM에 비해 효과적이면서 빠르게 입력 얼굴을 추적(Fitting)할 수 있으며 입력 영상의 정확한 3차원 얼굴 형태를 생성할 수 있음을 보였다.

키워드 : 3D 얼굴 인식, 3D 얼굴 추적, 얼굴 모델링, 3D 얼굴 형태

A 3D Face Reconstruction and Tracking Method using the Estimated Depth Information

Myung-Ho Ju[†] · Hang-Bong Kang[††]

ABSTRACT

A 3D face shape derived from 2D images may be useful in many applications, such as face recognition, face synthesis and human computer interaction. To do this, we develop a fast 3D Active Appearance Model (3D-AAM) method using depth estimation. The training images include specific 3D face poses which are extremely different from one another. The landmark's depth information of landmarks is estimated from the training image sequence by using the approximated Jacobian matrix. It is added at the test phase to deal with the 3D pose variations of the input face. Our experimental results show that the proposed method can efficiently fit the face shape, including the variations of facial expressions and 3D pose variations, better than the typical AAM, and can estimate accurate 3D face shape from images.

Keywords : 3D AAM, Face Modeling, 3D Face Shape

1. 서 론

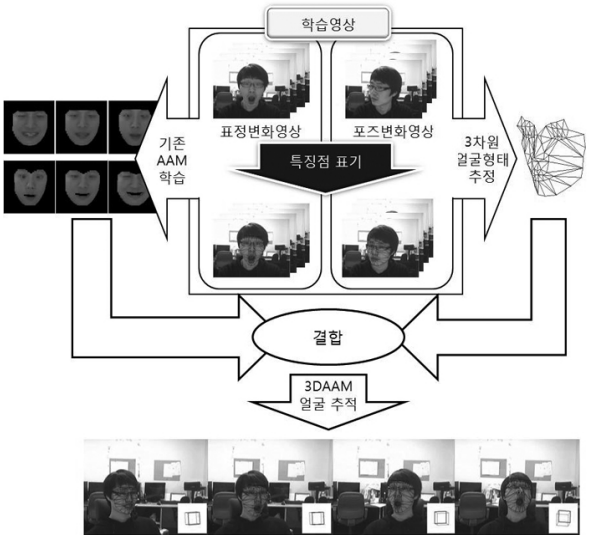
최근 비디오 영상에서 얼굴을 탐지하고 추적, 인식하기 위한 많은 연구들이 진행되어 왔다. 실제로 2차원 영상으로부터 3차원의 얼굴 형태를 추정하는 것은 얼굴 인식이나 표정 인식, 영상 모델링 등 다양한 컴퓨터 비전 분야에서 매우 유용하며 중요하다. 그러나 일반적으로 사용되는 카메라로부터 획득 가능한 영상은 영상 내 물체에 대한 깊이 정보를 포함하지 않는 2차원의 영상이기 때문에 이러한 영상으로부터 3차원 얼굴 형태를 추정하는 것은 매우 어려운 일이다.

얼굴의 3차원 정보 추정을 위해 가장 많이 이용되는 알고리즘은 3D Morphable Model (3DMM) [1] 이다. 3DMM은 얼굴의 형태와 질감 정보를 통계적으로 학습하고 이용하는 방법으로 효과적으로 얼굴의 3차원 정보를 표현할 수 있지

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2010년도 문화콘텐츠 산업기술지원사업의 지원 및 2010년도 가톨릭대학교 교비연구비 지원으로 이루어졌음.
† 준 회 원 : 가톨릭대학교 컴퓨터공학과 박사과정
†† 종신회원 : 가톨릭대학교 디지털미디어학부 교수(교신저자)
논문접수 : 2010년 9월 6일
수 정 일 : 1차 2010년 10월 19일, 2차 2010년 11월 5일
심사완료 : 2010년 11월 16일

만 3차원 스캐너로부터 획득된 정보를 직접 이용하여 학습하기 때문에 학습에 대한 요구 비용이 높다[2]. 이와 다르게 3차원의 데이터를 이용하지 않고 2차원 영상으로부터 효과적으로 3차원 형태를 추정하기 위한 방법으로 스테레오(Stereo) 기법[3]을 이용한 방법들이 제안되었다. Sung과 Kim [4]은 얼굴의 3차원 모델을 추정하기 위해 기존의 Active Appearance Model (AAM) 방법과 스테레오 기법을 결합한 2D+3D AAM을 제안하였다. 이 방법은 카메라로부터 입력된 두 장의 서로 다른 뷰(view)의 입력 얼굴 영상으로부터 스테레오 기법을 이용하여 깊이 정보를 추정하고 추정된 깊이 정보를 기존의 AAM[5]과 결합함으로써 3차원의 얼굴을 표현하였다. 그러나 스테레오 기법을 이용하여 깊이 정보를 추정하기 위해서는 많은 계산량을 요구하기 때문에 이 방법은 실시간(real-time)이나 근실시간(near real-time)으로 구현되기 어렵다. 또한 눈이나 입 등을 제외한 대부분의 얼굴 영역은 비슷한 색(피부 톤)을 가지기 때문에 스테레오를 통해 획득된 얼굴의 깊이 정보는 항상 정확한 깊이 정보로 추정되기 어렵다. Chen과 Wang [6]은 [4]와 유사하게 두 장의 스테레오 영상으로부터 깊이 정보를 추정하고 추정된 깊이 정보를 AAM학습에 이용되는 특징점(landmark)의 깊이 정보로 추가하여 3차원 얼굴 형태를 추정하는 3D AAM 방법을 제안하였다. 이 방법은 3DMM과 유사한 방법으로 얼굴 모델을 학습하지만 3차원의 데이터를 이용하지 않고 스테레오를 이용하여 3차원의 정보를 획득한다는 차이점을 갖는다. 그러나 앞서 언급한바와 같이 스테레오 기법을 이용하여 깊이 정보를 추정하는 것은 많은 비용을 요구하며 학습된 3D AAM을 이용하여 입력 얼굴의 3차원 형태를 추정하기 위해 기존의 AAM이 학습된 Jacobian 행렬을 이용하는 반면 그들의 방법에서는 최적화 단계마다 Jacobian 행렬을 새로 계산하기 때문에 입력 영상으로부터 빠르게 3차원 얼굴 형태를 추정하기 어렵다. 또한 그들의 연구는 3차원 얼굴 포즈의 변화에 의해 발생되는 self-occulsion으로 인해 좁은 각도의 얼굴 포즈 변화만을 고려하였다.

2차원 영상으로부터 3차원 정보를 추정하기 위한 일부 방법들[7, 8]은 3차원의 데이터를 직접 이용하는 3DMM과 기존의 AAM을 수학적으로 결합한다. 실제로 기존의 AAM과 3DMM 방법은 얼굴 형태의 추정 과정에서 많은 유사성을 갖는다. 단지 서로 다른 차원을 이용하기 때문에 기존의 AAM은 2차원의 얼굴 형태를, 3DMM은 3차원의 얼굴 형태를 추정한다. Xiao와 Baker [7]는 이러한 유사성을 이용하여 2차원의 얼굴 정보를 분해하고 분해된 결과로써 3차원의 얼굴 형태를 추정하였다. 2차원의 얼굴 형태는 3차원의 얼굴로부터 임의의 투영행렬에 의해 투영되었다고 가정하고 Singular Value Decomposition (SVD)을 이용하여 2차원 학습 얼굴 특징점의 데이터를 3차원의 데이터와 투영행렬로 분해하였다. Heo와 Savvides [8]은 [7]의 방법으로부터 획득된 3차원 얼굴 형태에서 특징점을 세분화 되도록 분해하여 보다 밀집된 3차원의 얼굴 형태를 추정하였다. 그러나 이들의 방법은 2차원의 얼굴 특징점들로부터 정확한 3차원의 얼굴 특징점 정보를 추정하기 위해서 매우 많은 수의 학습 영상을 요구한다. 학습에 사용되는 영상의 수는 테스트 단계에 영향을 주지 않지만 실제 응용되는 프로그램에 따라 적은 수의 학습 영상만이 주어질 경우, 제안된 방법은 효율적으로 동작되기 어렵다.



(그림 1) 제안된 시스템의 학습 모듈

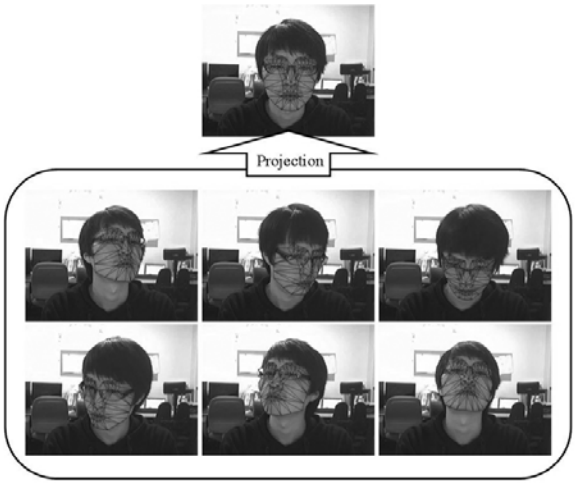
본 논문에서는 얼굴의 깊이 정보 추정을 이용하는 빠른 3D-AAM 방법을 제안한다. 제안된 방법은 기존 AAM의 빠른 얼굴 추적(Fitting) 능력을 유지하면서 입력 얼굴의 3차원 변화를 효과적으로 추정, 해당 3차원 얼굴 모델을 생성한다. 얼굴의 각 특징점의 깊이 정보는 서로 다른 얼굴 포즈를 갖는 학습 영상간의 기하학적인 변화를 이용하여 추정할 수 있다. 추정된 각 특징점의 깊이 정보는 기존의 AAM 방법과 결합하여 빠르게 2차원 영상에서 3차원 얼굴 형태 정보를 추정한다. 제안되는 방법은 얼굴 형태를 추정하는 최적화 단계에서 [6]과 다르게 학습된 Jacobian을 그대로 이용하기 때문에 기존의 AAM과 유사하게 임의의 3차원 얼굴 포즈 변화와 표정 변화를 갖는 입력 얼굴 영상을 매우 빠르게 추정할 수 있다. (그림 1)은 제안된 시스템의 학습 모듈을 보여준다. 얼굴 표정 변화와 포즈 변화를 가지는 두 가지의 학습 영상으로부터 각각 기존의 AAM 학습과 3차원 얼굴 모델 추정을 수행한다. 본 논문에서는 학습된 AAM과 추정된 3차원 얼굴 모델을 효율적으로 결합함으로써 다양한 표정 변화와 함께 3차원의 얼굴 포즈 변화를 포함한 입력 영상을 빠르게 추적(Fitting)하고 입력 영상의 정확한 3차원 얼굴 형태를 생성한다.

본 논문의 구성은 다음과 같다. 2절에서는 서로 다른 얼굴 포즈를 갖는 학습 영상들로부터 얼굴의 각 특징점의 깊

이 정보를 추정하기 위해 제안되는 방법을 설명한다. 3절에서는 2절에서 추정된 각 특징점의 깊이 정보를 기존의 AAM에 결합하여 임의의 3차원 얼굴 변화가 표현되는 2차원 영상으로부터 3차원 얼굴 형태를 추정하는 3D-AAM 방법을 설명한다. 그리고 4절에서 실험결과를 통해 본 논문에서 제시하는 방법이 기존의 연구보다 우수한 성능을 가짐을 보이며 5절에서 결론을 맺는다.

2. 3차원 깊이 추정

입력 얼굴 영상으로부터 깊이 정보를 추정하기 위해 본 논문에서는 먼저 각 사람으로부터 학습 영상을 생성한다. 학습 영상은 (그림 2)와 같이 무표정의 얼굴로 상, 하, 좌, 우 방향의 3차원 얼굴 포즈 변화를 포함한다. 각 학습 영상은 수공(Manual)으로 얼굴의 특징점을 표기하고 3차원 얼굴 포즈 방향에 대한 대략적인 회전 각도를 할당한다. 이때, 할당된 특징점은 2차원의 학습 영상에 표기되기 때문에 x와 y의 2차원 정보만을 가지며 할당된 3차원의 얼굴 포즈 각도는 대략적으로 할당되기 때문에 정확한 3차원 정보를 나타내지 않는다.

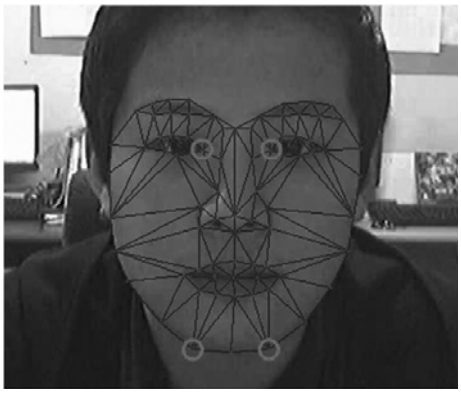


(그림 2) 6장의 서로 다른 얼굴 포즈 영상을 이용한 깊이 정보 추정 알고리즘

각 특징점은 모든 학습 영상에서 동일한 얼굴 위치에 표기되기 때문에 특징점을 학습 영상간의 매칭점으로 고려할 수 있다. 그러므로 임의의 두 학습 영상간의 관계는 다음과 같이 정의될 수 있다.

$$\mathbf{X}_i^k = \mathbf{H}_{ij}\mathbf{X}_j^k \tag{1}$$

여기서 $\mathbf{X}_i^k$는 i번째 학습 영상의 k번째 특징점의 동치좌표(homogeneous coordinates)로 $\mathbf{X}_i^k = [x, y, z]$이다. 특징점은 2차원의 정보로 표현되기 때문에 깊이 좌표 z는 최초에 0으로 설정한다. 그리고 $\mathbf{H}_{ij}$는 i번째 학습 영상과 j번째 학습 영상간의 Homography로 다음과 같이 정의된다.



(그림 3) 얼굴의 매쉬(Mesh): 붉은색 원의 특징점은 Homography를 계산하기 위해 사용된다.

$$\mathbf{H}_{ij} = \mathbf{T}_i\mathbf{R}_i\mathbf{R}_j^T\mathbf{T}_j^{-1} \tag{2}$$

여기서 $\mathbf{T}_i$와 $\mathbf{R}_i$는 각각 i번째 학습 영상의 크기 변화와 이동 변화를 나타내는 변환 행렬(Transformation matrix)과 영상의 회전 행렬(Rotation matrix)을 나타낸다.

일반적으로 두 영상간의 Homography는 RANSAC[9]을 이용하여 계산된다. 그리고 Homography를 계산하기 위해서는 최소한 4개의 매칭점이 필요하다. 그러나 얼굴의 특징점은 서로 다른 깊이 정보를 가지기 때문에 RANSAC을 이용하여 두 영상간의 Homography를 계산할 경우, 선택된 4개의 특징점이 항상 같은 평면에 위치하기 어렵기 때문에 정확한 Homography를 계산하기 어렵다. 그러므로 본 논문에서는 정확한 Homography를 계산하기 위해 (그림 3)과 같이 동일한 평면 위에 위치하는 4개의 사전 정의된 특징점을 이용하여 영상간의 Homography를 계산한다.

Homography가 계산된 후에 식(2)의 회전행렬, $\mathbf{R}_i$은 대략적으로 할당된 학습 영상의 얼굴 포즈 각도로 할당한다. 그리고 계산된 Homography와 회전행렬을 이용하여 변환 행렬, $\mathbf{T}_i$를 계산할 수 있다.

현재 값을 가지고 있지 않거나 정확한 값을 가지지 않은 학습 영상의 각 특징점의 깊이 좌표, z, 회전 행렬 파라미터 $[\theta_x, \theta_y, \theta_z]$의 정확한 값을 계산하기 위해 본 논문에서는 최적화 알고리즘의 하나인 번들 과정(Bundle adjustment phase)[9]을 통해 반복적으로 각 값을 갱신한다. 번들 과정을 위해 사용되는 목적 함수(Objective function)는 제곱 투영 에러(squared projection error)의 합을 이용하였다. 이는 학습 영상의 모든 특징점이 평균 형태의 정면 얼굴 영상으로 투영되었을 경우의 특징점간의 에러를 나타낸다. 번들

과정에서 각 특징점의 깊이 좌표, z와 회전 파라미터에 따라 제곱 투영 에러의 합을 최소화 한다. i번째 학습 영상의 k번째 특징점의 잔여 에러, r_i^k는 다음과 같이 정의된다.

$$r_i^k = \mathbf{X}^k - \mathbf{P}_i^k \tag{3}$$

여기서 $\mathbf{X}^k$는 투영되는 정면 얼굴 영상의 k번째 특징점의 좌표이며 $\mathbf{P}_i^k$는 i번째 영상에서 정면 얼굴 영상으로 투영되는 k번째 특징점의 투영 좌표로 다음과 같이 표현된다.

$$P_i^k = T_1 R_1 R_i^T T_i^{-1} X_i^k \tag{4}$$

에러 함수는 모든 학습 영상에서 정면 얼굴 영상으로 투영되는 잔여 에러의 합으로 다음과 같이 정의된다.

$$e = \sum_{i=2}^{N} \sum_{k=1}^{M} |r_i^k|^2 \tag{5}$$

여기서 N은 학습 영상의 개수이며 M은 얼굴에 표기된 특징점의 개수이다. 그리고 학습 영상의 첫 번째 영상을 모든 학습 영상이 투영되는 정면 얼굴 영상으로 가정하였다.

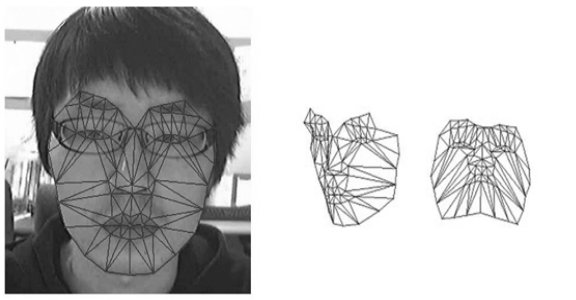
위와 같은 최소화 문제는 비선형 최소 제곱 문제(non-linear least squares problem)로 본 논문에서는 최소화를 위해 Levenberg-Marquardt 알고리즘[9]을 이용하였다. 최소화를 위한 각 단계는 다음과 같이 계산된다.

$$\boldsymbol{\Phi} = \left(\mathbf{J}^T\mathbf{J} + \lambda\mathbf{C}^{-1}\right)^{-1}\mathbf{J}^T\mathbf{r} \tag{6}$$

여기서 $\boldsymbol{\Phi} = \left\{z_1, \ldots, z_M, \theta_x^1, \theta_y^1, \theta_z^1, \ldots, \theta_z^N\right\}$로 갱신되는 모든 파라미터를 나타내고 $\mathbf{r}$은 식(3)의 잔여 에러이다. 그리고 $\mathbf{J}$는 Jacobian 행렬이며 $\mathbf{C}$는 각 파라미터 변화에 대한 사전 정보(prior belief)를 포함하는 공분산 행렬(Covariance matrix)로 정의된다.

3. 3D Active Appearance Model

얼굴의 형태는 (그림 4(a))와 같이 M개의 2차원 특징점으로 표현된다. 많은 기존의 방법에서 얼굴의 특징점은 에지(Edge)나 모서리(Coner)와 같은 주요한 위치(salient location)에 표기한다. 이와 더불어 본 논문에서는 얼굴의 3차원 정보를 표현하기 위해 코끝이나 이마의 중점과 같은 3차원의 모서리에 추가적으로 특징점을 표기하였다. 추가된 특징점은 3차원 얼굴 포즈 변화로 인해 발생되는 Self-occusion으로 매쉬가 가려지는 것을 방지한다.



(a) The mesh for 3D AAM (b) 3D pose variations of the mesh

(그림 4) 제안된 방법의 3차원 변환 얼굴 매쉬(Mesh)

얼굴 형태의 변화를 학습하기 위해 AAM은 먼저 Procustes analysis[10] 방법을 이용하여 학습 영상으로부터 추출된 얼굴 형태를 얼굴의 평균 형태, $\bar{s}$로 반복적으로 정렬한다. 그러나 본 논문에서는 얼굴의 깊이 정보를 추정하기 위해 무표정의 정면 얼굴을 이용하기 때문에 이러한 정면 얼굴을 평균 얼굴로 가정하고 학습 영상을 정면 얼굴에 정렬하였다.

각 학습 얼굴의 형태는 $M \times 2$개의 원소를 갖는 벡터로 정의 되며 얼굴 형태 공간은 평균 얼굴 형태 벡터를 중심으로 학습 얼굴 형태 벡터들에 Principal Component Analysis(PCA)를 적용하여 생성한다. 형태 모델, S_M은 형태 공간으로부터 고유 행렬, $p = \{p_1, \ldots, p_l\}$의 가중치와 형태 파라미터, $s = \{s_1, \ldots, s_l\}$의 선형 결합으로 다음과 같이 표현된다.

$$S_M = \bar{s} + \sum_{i=1}^{l} p_i s_i \tag{7}$$

얼굴의 외관 공간(Appearance Subspace)은 각 학습 얼굴의 얼굴 형태 변화로 인한 질감의 변화를 제거하기 위해 평균 얼굴 형태로 와핑(Warping)되어 형태 자유 영상(shape-free-image)으로 변환되고 얼굴 형태와 같이 벡터의 집합으로 표현된다. 외관 모델, G_M은 평균 얼굴 외관 벡터를 중심으로 학습 외관 벡터 집합에 PCA를 적용하여 고유 행렬, $g = \{g_1, \ldots, g_m\}$의 가중치와 외관 파라미터, $A = \{A_1, \ldots, A_m\}$의 선형 결합으로 표현된다.

$$G_M = \overline{A} + \sum_{i=1}^{m} g_i A_i \tag{8}$$

AAM은 형태 파라미터와 외관 파라미터를 연결하여 PCA를 수행함으로써 형태 모델과 외관 모델을 결합한다.

$$[\kappa S_M, G_M]^T = \sum_{i=1}^{n} q_i c_i \quad (9)$$

여기서 κ는 형태 파라미터의 가중치로 형태와 외관간의 차이를 최소화 한다. 본 논문에서는 κ를 형태 변환의 총합과 외관의 픽셀 변화의 총합의 비율로 설정하였다. (그림 5)는 첫 번째 AAM 파라미터가 변할 경우의 얼굴 형태와 외관 변화의 예를 보여준다.

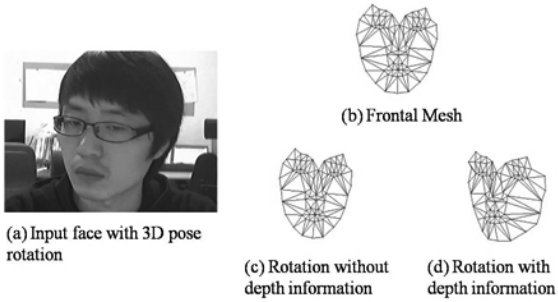
(그림 5) 첫 번째 AAM 파라미터를 변화시켰을 경우의 얼굴 형태 및 외관 예: 중앙 그림은 평균 AAM모델이며 좌, 우 영상은 첫 번째 파라미터를 감소시켰을 경우와 증가시켰을 경우의 결과 영상이다.

본 논문에서 입력 얼굴에 따른 임의의 3차원 얼굴 형태는 3차원 얼굴 좌표에 해당되는 투영 변환, $\mathbf{X} = S_t(\mathbf{X})$을 적용하여 생성된다고 가정한다[11]. 여기서 S_t는 3차원 회전과 이동을 표현하는 변환으로 다음과 같이 정의된다.

$$S_t(\mathbf{X}) = \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{X} \quad (10)$$

여기서 $\mathbf{X}$는 AAM으로부터 생성된 2차원 얼굴 특징점에 2절에서 추정된 깊이 정보를 추가한 3차원 특징점의 집합이다. 그리고 $\mathbf{T}$와 $\mathbf{R}$은 3차원 회전 행렬과 이동 행렬을 나타낸다. Rodriguez Equation[9]를 이용하여 회전 행렬은 3개의 회전 파라미터, $[\theta_x, \theta_y, \theta_z]$로 표현될 수 있으며 이동 행렬은 3개의 이동 파라미터, $[t_x, t_y, t_z]$로 표현된다. 3차원 이동 파라미터는 z축에 따른 이동으로 얼굴의 크기 변화를 포함한다. 그러므로 얼굴의 3차원 포즈 파라미터 벡터는 $\mathbf{t} = (\theta_x, \theta_y, \theta_z, t_x, t_y, t_z)^T$으로 표현되며 가장 이상적인 경우, 모든 파라미터가 0의 값을 갖는다.

일반적으로 얼굴의 포즈나 표정이 변할 때, 얼굴 특징점의 깊이 정보 역시 변화한다. 그러나 얼굴 포즈 변화 없이 얼굴의 표정만 변화할 경우 각 특징점의 깊이 정보 변화는 매우 작다. 그러므로 본 논문에서는 포즈 변화 없이 얼굴의 표정만이 변할 경우, 얼굴의 깊이 정보는 변하지 않는다고 가정한다. 이러한 가정으로 본 논문에서는 AAM 모델 파라미터, $\mathbf{c}$에 의해 생성된 2차원의 얼굴 모델에 추정된 각 특징점의 깊이 정보를 z축으로 추가함으로써 얼굴의 3차원 형태를 생성한다. 생성된 3차원 얼굴 형태에 3차원 변환 파라미터, $\mathbf{t}$를 적용하고 영상에 평형 투영(parallel projection) 함으로서 최종 얼굴 영상을 생성한다. 그러므로 제안된 방법은 임의의 얼굴 포즈 및 표정을 갖는 영상으로부터 다음 파라마터, $\mathbf{p}$를 추정한다.



(그림 6) 깊이 정보를 포함한 매쉬와 포함하지 않은 매쉬의 3차원 회전 변환 예

$$\mathbf{p}^T = (\mathbf{c}^T | \mathbf{t}^T) \quad (11)$$

AAM 모델 파라미터, $\mathbf{c}$에 의해 생성된 얼굴 형태는 (그림 6(b))와 같이 2차원의 형태를 갖는다. 그러므로 입력 얼굴에 따른 임의의 3차원 변환, $\mathbf{t}$가 주어졌을 때, (그림 6(c))와 같은 변환을 보인다. 그러나 본 논문에서 제안된 방법은 (그림 6(d))와 같이 정확한 3차원 모델을 생성한다.

입력 얼굴에 맞는 3D-AAM 모델을 생성하기 위해 제안된 방법은 입력 영상으로부터 투영된 질감 모델, G_S와 3D-AAM으로부터 생성된 합성 모델, G_M간의 차이를 파라미터 벡터, $\mathbf{p}$에 따라 최소화한다.

$$r(\mathbf{p}) = G_S - G_M \quad (12)$$

식(12)의 일차 테일러 급수(first order Taylor expansion)는 다음과 같다.

$$r(\mathbf{p} + \delta\mathbf{p}) = r(\mathbf{p}) + \frac{\partial r}{\partial \mathbf{p}} \delta\mathbf{p} \quad (13)$$

시스템은 잔여 에러, r이 주어졌을 때, $|r(\mathbf{p} + \delta\mathbf{p})|^2$을 최소화하는 $\delta\mathbf{p}$를 선택한다. 그러므로 식(13)을 0으로 정의하여 다음과 같은 RMS 식을 얻을 수 있다.

$$\delta\mathbf{p} = -\mathbf{R} r(\mathbf{p}) \quad where \quad \mathbf{R} = \left(\frac{\partial r}{\partial \mathbf{p}}^T \frac{\partial r}{\partial \mathbf{p}} \right)^{-1} \frac{\partial r}{\partial \mathbf{p}}^T \quad (14)$$

기존의 AAM과 유사하게 본 논문에서는 $\delta\mathbf{p}$를 계산하기 위한 행렬, $\mathbf{R}$을 학습 과정에서 추정한다. 그러나 기존

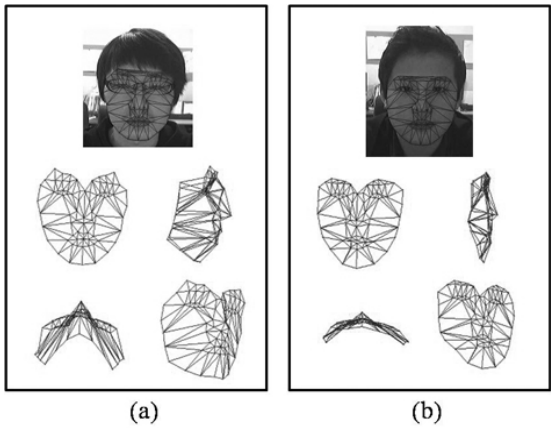
AAM과 다르게 $\mathbf{R}$을 계산하기 위한 Jacobian 행렬, $\mathbf{J}=\frac{\partial r}{\partial \mathbf{p}}$은 추정된 깊이 정보를 이용하는 3차원 변환을 포함하는 식(11)의 파라미터 벡터, $\mathbf{p}$에 의해 계산된다.

4. 실험 결과

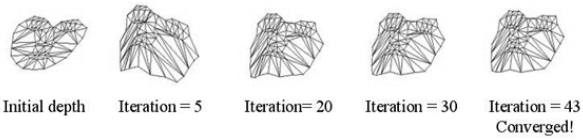
입력 얼굴 영상의 3차원 변화를 학습하기 위해 본 논문에서는 5명의 사용자로부터 두 가지 종류의 학습 영상을 생성하였다. 생성된 모든 학습 영상은 약 15초의 길이와 15 fps를 가진다. 각 사용자의 첫 번째 학습 영상은 정면 얼굴을 유지하면서 다양한 얼굴 표정 변화를 포함한다. 학습 영상이 총 225개의 프레임으로 구성되지만 본 논문에서는 10개 프레임 간격으로 22개의 학습 영상을 선택하여 얼굴 표정 변화 학습에 사용하였다. 두 번째 학습 영상은 얼굴의 표정 변화 없이 다양한 3차원 얼굴 포즈 변화를 포함한다. 본 논문에서는 (그림 2)와 같이 정면 포즈의 얼굴과 서로 크가 다른 6개 이상의 영상을 선택하여 얼굴의 깊이 정보 추정을 위해 사용하였다. 제안된 3D-AAM을 학습하기 위해 먼저 각 선택된 학습 영상에 수공으로 (그림 4)와 같이 80개의 얼굴 특징점을 표기하였다.

(그림 7)은 제안된 방법의 깊이 정보 추정을 통한 3차원 모델 생성의 예를 보여준다. (그림 7(a))는 약간 마른 사용자로부터 추정된 3차원 얼굴 형태 매쉬이며 (그림 7(b))는 보다 통통한 사람으로 둥근 얼굴형의 사용자로부터 추정된 3차원 얼굴 형태 매쉬를 보여준다. 그림은 3차원 형태 매쉬의 정면과 측면, 하단, 대각측면의 매쉬 영상을 보여준다.

얼굴 특징점의 깊이 정보를 추정하여 3차원 얼굴 형태를 생성하기 위해 Xiao[7]는 180장의 매우 많은 학습 영상을 이용하였다. 그러나 본 논문에서 제안된 방법은 서로 크게 다른 최소 6장의 얼굴 영상을 이용여 효율적으로 얼굴의 3차원 형태를 추정할 수 있었다. (그림 2)는 제안된 방법에 사용된 6장의 서로 다른 얼굴 포즈 영상의 예를 보여준다. 그리고 (그림 8)은 제안된 방법에서 번들 과정을 통해 반복적으로 갱신되는 3차원 얼굴 형태의 예이다. 6장의 얼굴 포즈 영상은 정면 얼굴 영상에 투영되어 식(5)의 잔여에러를 갖는다. 서로 다른 얼굴 포즈 영상의 정면 얼굴 영상에 대한 투영은 3차원 투영이기 때문에, 잔여 에러는 각 landmark의 깊이 정보가 올바르지 않을 경우, 높은 에러를 가지며 정확한 깊이 정보를 가질수록 정면 얼굴 영상에 올바르게 투영되어 낮은 잔여 에러를 갖는다. 본 논문에서는 최적화(Optimization)방법의 하나인 번들과정[9]을 이용하여 입력 영상의 잔여에러를 반복적으로 최소화함으로써 최종적으로 정확한 3차원 얼굴 깊이 정보를 추정한다. 번들 과정은 식(6)과 같이 투영 에러에 따른 변환을 나타나는 Jacobian행렬을 이용하여 영상의 투영 에러가 주어졌을 때,



(그림 7) 깊이 추정 결과 예: (a) 마른 사람의 3차원 얼굴 생성 예, (b) 살찐 사람의 3차원 얼굴 생성 예



(그림 8) 제안된 방법의 깊이 추정: 입력 영상의 깊이 정보는 번들 과정을 통해 반복적으로 갱신되며 최종적으로 최적화된다.

잔여에러를 최소화하는 3차원 얼굴 모델의 깊이 정보와 6장의 포즈 영상의 3차원 얼굴 회전각도 파라미터 변화량을 추정한다. 그리고 이를 이용하여 파라미터를 반복적으로 갱신함으로써 잔여 에러를 최소화하여 최종적으로 (그림 8)과 같이 올바른 3차원 얼굴 형태를 추정한다. 본 논문에서는 얼굴 포즈가 크게 다른 6개의 영상을 이용하여 3차원 얼굴 형태를 추정하였다. 번들 과정에서는 약 10번의 반복 과정에서 입력된 얼굴 형태와 매우 유사한 3차원 얼굴 형태가 추정되었으며 최종적으로 43번의 반복 후에 가장 작은 잔여에러를 갖는 3차원 얼굴 모델로 최적화(Converge)되었다.

본 논문에서는 제안된 방법의 성능을 평가하기 위해 기존의 AAM[5]과 정확성과 수행 속도 측면을 비교하였다. 테스트 영상은 학습 영상과 별도로 제작되었으며 학습 영상과 동일한 길이를 갖지만 사용자가 원하는 임의의 다양한 얼굴 표정과 자연스러운 3차원 얼굴 포즈 변화를 포함하였다. (그림 9)는 테스트 영상에서 얼굴 추적 및 3차원 얼굴 형태 모델 생성 결과의 주요 프레임을 보여준다. (그림 9(a))와 같이 본 논문에서 제안된 방법은 임의의 얼굴 표정 및 3차원 얼굴 포즈에 대해 정확한 얼굴 추적 및 3차원 형태 모델을 생성하는 반면, (그림 9(b))와 같이 기존의 AAM에서는 입력 얼굴을 올바르게 추적하지 못하거나 얼굴의 위치를 놓치게 되는 것을 볼 수 있었다. 또한 제안된 방법은 3차원의 얼굴 형태를 생성하기 때문에 각 영상 프레임에서의 얼굴 포즈의 방향 각도를 추정할 수 있었다. (그림 9(a))의 직육면체는 얼굴의 포즈 방향을 나타내며 붉은색 면은 얼굴의 정면 방향을 나타낸다.

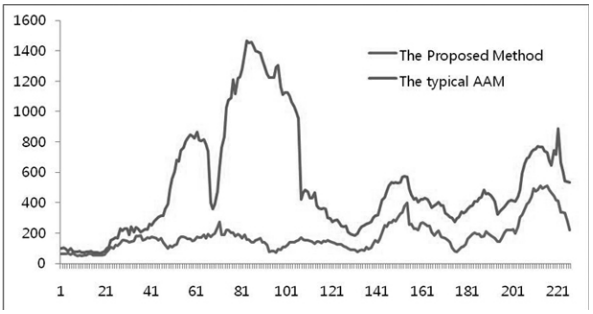
(a) Some fitting results of our method including the estimated 3D rotation angle.

(b) Some fitting results of the typical AAM method.

(그림 9) 제안된 방법과 기존의 AAM 방법의 얼굴 추적 결과 비교

(그림 10)은 테스트 영상 중 한명의 사용자에 대한 택스처 매칭의 평균 RMS 에러의 변화이다. 텍스처 매칭 에러는 입력 얼굴과 생성된 얼굴 모델의 텍스처 차이로 유사할 경우, 낮은 에러를 갖지만 시스템이 얼굴을 올바르게 추적하지 못하거나 놓칠 경우 높은 에러를 가지므로 얼굴 추적 성능을 평가하는데 매우 용이하다. (그림 10)과 같이 본 논문에서 제안된 방법은 임의의 얼굴 표정과 얼굴 포즈의 입력 얼굴에 대해 올바르게 얼굴을 추적하고 모델을 생성하기 때문에 입력 얼굴을 올바르게 추적하지 못하는 기존의 AAM에 비해 매우 낮은 에러를 보였다. (그림 10)의 40 ~100 프레임에서 기존의 AAM은 입력 얼굴의 위치를 놓친 경우가 발생하여 매우 높은 에러를 나타냈다.

본 논문에서 제안된 방법과 기존의 AAM[5], 스테레오를 이용한 2D+3D AAM[4], 3DAAM[6]은 모두 Visual C++ 2005의 프레임 워크에서 C 기반으로 구현되었으며 Intel Quad Q8200 2.33 GHz와 4GB의 메모리를 갖는 컴퓨터에서 테스트 되었다. AAM을 이용하여 입력 얼굴의 형태를 추정(Fitting)하는 것은 얼굴 형태를 표현하는 파라미터를 추정하는 것과 동일하다. 이러한 파라미터의 추정은 최적화(Optimization) 알고리즘의 하나로 반복적으로 파라미터에 대한 에러를 최소화하도록 각 파라미터를 갱신한다. <표 1>은 기존의 AAM[5]과 본 논문과 유사한 3차원 얼굴 파라미터를 사용하는 2D+3D AAM[4], 3DAAM[6], 그리고 본 논문에서 제시한 방법에 대한 입력 얼굴 피팅(fitting)을 위한



(그림 10) 한 사람의 테스트 영상에 대한 평균 RMS 텍스처 매칭 에러 변화율

각 반복 단계에서의 평균 수행 시간을 나타낸다. 평균 수행 시간은 10,000번의 수행 결과를 평균하였다. 기존의 AAM의 경우, 2차원의 얼굴 형태를 표현하는 파라미터를 추정하기 때문에 가장 짧은 수행 시간을 보인다. 이에 반해 기존의 2D+3D AAM과 3DAAM, 그리고 본 논문에서 제시하는 방법은 얼굴의 3차원 정보를 표현하기 위해 기존 AAM에 비해 보다 많은 파라미터를 추정해야 하기 때문에 추가적인 수행 시간을 요구한다. 3차원 얼굴 파라미터를 추정하는 기존의 2D+3D AAM과 3DAAM에 비해 본 논문에서 제안된 방법은 얼굴의 3차원 포즈 파라미터를 추가적으로 가지기 때문에 가장 많은 수행 시간을 필요로 한다. 그러나 본 논문에서 제안된 방법은 기존 방법들에 비해 다양한 얼굴 표정과 함께 3차원 얼굴 포즈 변화에 대해 올바르게 입력 얼굴을 추적하고 3차원 얼굴 모델을 생성하면서도 기존의 AAM에 비해 0.21 ms, 기존의 2D+3D AAM, 3DAAM에 비해 약 0.005ms의 매우 짧은 수행 시간만을 더 요구하였다. 또한 본 논문에서 제안된 방법은 기존의 AAM과 같이 입력 얼굴에 대해 파라미터 추정을 위한 피팅 시간만을 요구하지만 얼굴의 3차원 정보가 추정 가능한 기존 2D+3D AAM[4]과 3DAAM[6]의 경우, 스테레오를 기반으로 하기 때문에 입력 얼굴의 피팅 수행과 함께 매 입력 영상마다 3차원 정보를 획득하기 위한 스테레오 깊이 추정 시간이 필요하다. 본 논문의 테스트 환경에서 입력 영상이 320×240 크기를 가지며 3×3 크기의 윈도우를 사용할 경우, 가장 간단한 SSD(Sum of Squared Distance) 방법은 약 4055 ms의 시간이 소요되었으며, ASW(Adaptive Support Weight)[3]의 방법을 이용할 경우, 8860 ms의 시간이 얼굴 피팅 시간 이외에 추가로 요구되었다.

〈표 1〉 각 반복 단계에서의 수행 시간 비교

방 법	기존의 AAM[5]	2D+3D AAM[4]	3DAAM[6]	제안된 방법
각 반복 단계에서의 수행 시간	0.521 ms	0.537 ms	0.539 ms	0.542 ms

5. 결 론

본 논문에서는 서로 얼굴 포즈가 크게 다른 학습 영상으로부터 얼굴 특징점간의 기하학적 변화를 이용하여 각 특징점의 깊이 정보를 추정하였다. 그리고 추정된 깊이 정보를 기존의 AAM과 결합하여 임의의 얼굴 표정 및 3차원 얼굴 포즈를 갖는 입력 영상에 대해 빠르게 얼굴 형태를 추적하고 3차원 얼굴 형태를 생성하기 위한 3D-AAM 방법을 제안하였다. 제안된 방법은 학습 영상의 얼굴 포즈가 크게 다를 경우, 최소 6장의 학습 영상만으로도 효과적인 3차원 얼굴 형태 모델을 생성하였다. 또한 학습 영상에서 추정된 Jacobian 행렬을 테스트 단계의 최적화 과정에 이용함으로써 입력 얼굴의 3차원 모델을 빠르게 생성하고 추적(Fitting)할 수 있었다. 제안된 방법은 2차원 입력 영상으로부터 3차원의 얼굴 형태를 생성하면서도 기존의 AAM에 비해 약간의 수행 시간만이 추가로 요구되었다.

그러나 얼굴의 포즈가 크게 변하여 Self-occlusion이 매우 크게 발생할 경우, 사용 가능한 얼굴 텍스처의 부족으로 올바르게 얼굴의 형태를 추적하지 못하였다. 그러므로 이러한 Self-occlusion 및 임의 물체의 occlusion에 대해 효과적인 얼굴 추적 및 3차원 모델 생성은 앞으로 나아가야 할 방향이라 하겠다.

참 고 문 헌

[1] N. Faggian, A. P. Paplinski, and J. Sherrah. "Active Appearance Models for Automatic Fitting of 3D Morphable Models". IEEE International Conference on Video and Signal Based Surveillance, 90, 2006.

[2] Skoglund. "Three-dimensional face modeling and analysis", M.S. thesis, Informatics and Mathematical Modelling, Tech. Univ. Denmark, Lyngby, Denmark, 2003.

[3] K.-J. Yoon, and I. S. Kweon. "Adaptive Support-Weight Approach for Correspondence Search", IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4):650-656, 2006.

[4] J. Sung, and D. Kim. "Pose-Robust Facial Expression Recognition Using View-Based 2D+3D AAM", IEEE Transactions On Systems, Man and Cybernetics, Part A: SYSTEMS AND HUMANS, 38(4):852-866, 2008.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active Appearance Models", IEEE Transactionis on Pattern Analysis and Machine Intelligence, 23(6):681-685, 2001.

[6] C.-W. Chen, and C.-C. Wang. "3D Active Appearance Model for Aligning Faces in 2D Images", Proceedings of the IEEE/RS International Conference on Intelligent Robots and Systems, 3133-3139, 2008.

[7] J. Xiao, S. Baker, I. Matthews, and T. Kanade. "Real-Time Combined 2D+3D Active Appearance Models", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 535-542, 2004.

[8] J. Heo, and M. Savvides. "In Between 3D Active Appearance Models and 3D Morphable Models", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 20-26, 2009.

[9] M. Brown, and D. G. Lowe. "Automatic Panoramic Image Stitching using Invariant Features", International Journal of Computer Vision, 74(1):59-73, 2007.

[10] M. D. Cordea, and E. M. Petriu. "A 3-D Anthropometric-Muscle-Based Active Appearance Model," IEEE Transactions on Instrumentation and Measurement, 55(1):91-98, 2006.

[11] R. Hartley, and A. Zisserman. "Multiple View Geometry in Computer Vision," Cambridge University Press, ISBN:0521540518, second edition, 2004.

주 명 호

e-mail : hangel5@catholic.ac.kr
2005년 가톨릭대학교 컴퓨터공학과(학사)
2007년 가톨릭대학교 컴퓨터공학과(석사)
2007년~현 재 가톨릭대학교 컴퓨터공학과(박사과정)
관심분야 : 영상처리, 인공지능, 컴퓨터비전

강 행 봉

e-mail : hbkang@catholic.ac.kr
1980년 한양대학교 전자공학과(학사)
1986년 한양대학교 전자공학과(석사)
1989년 Ohio State Univ. 컴퓨터공학(석사)
1993년 Rensselaer Polytechnic Institute 컴퓨터 공학(박사)
1993년~1997년 삼성종합기술원 수석연구원
1997년~현 재 가톨릭대학교 디지털미디어학부 교수
2005년 UC Santa Barbara, Visiting Professor
관심분야 : 컴퓨터비전, HCI, 컴퓨터그래픽스, 인공지능